Photorealistic Style Transfer for Cinema Shoots

Itziar Zabaleta
Dpt. of Information and Communication Technologies
Universitat Pompeu Fabra
Barcelona, Spain
itziar.zabaleta@upf.edu

Marcelo Bertalmío
Dpt. of Information and Communication Technologies
Universitat Pompeu Fabra
Barcelona, Spain
marcelo.bertalmio@upf.edu

Abstract—**Color grading is the process of subtly mixing and adjusting the color and tonal balance of a movie to achieve a specific visual *look*. This manual editing task may require a significant amount of work from very skilled artists and technicians. In many cases the director wants to emulate the style and look present in a reference image, e.g. a still from an existing movie, a photograph, or even a previously shot sequence in the current movie. In this paper we propose a method that automatically transfers the style, in terms of tone, color palette and contrast, from a reference image to the source RAW image. It consists of three separable operations: global luminance matching, global color transfer and local contrast matching. The computational complexity of the procedure is extremely low and allows for real-time implementation in-camera. As it just takes into account the statistics of source and reference images, no training is required. The results are free from artifacts and provide an excellent approximation to the intended look, bringing savings in pre-production, shooting and post-production time.**

Index Terms—**Image processing, computational photography, style transfer, color grading, visual look.**

I. INTRODUCTION

Post-production is the part of the process of filmmaking, video production, and photography that includes all stages of production occurring after shooting. Color grading is one of the last steps in post-production that enables to elevate the visual appeal of the images and achieve a certain look, relative to the narrative and artistic needs of a program. Color is an integral part of film, and it plays a critical role in how we perceive the film and the characters. It can build harmony or tension within a scene, or bring attention to a key themes. When chosen deliberately, a well-placed movie color palette evokes mood and sets the tone for the film.

Ideally, the intended color interaction would be conceived on the set, recorded faithfully, and enhanced carefully by the colorist under strict supervision. In the past color grading was a photo-chemical process performed at a photographic laboratory, but nowadays is generally performed digitally in a color suite. This process is very costly in terms of budget and time and it may require a significant amount of work from very skilled artists and technicians.

We propose a method that can ease the workload of cinematographers and photographers during post-production. In many cases the director wants to emulate the style and look present in a reference image, e.g. a still from an existing movie, a photograph, or even a previously shot sequence in the current movie. Given a reference image, our approach automatically transfers the style, in terms of tone, color palette and contrast to the source RAW image. It is a real-time process that can be implemented in-camera, so light and color can be adjusted on-set while seeing the resulting image on the screen. While the method is proposed as a substitute for some of the post-production tasks, it is compatible with further refinements, both on-set and in post-production. The method is applied directly on the RAW images and it generates a display-ready result that matches the style of the reference image.

There is extensive work on color transfer between images, some of them are low-complexity methods and can work real-time [8], [12], [14]. However, they work with non-linearized images and not directly with RAW data. Instead, our approach takes as input RAW data that is non-linearly transformed in the first step of the method. An automatic method to transfer retouching styles to RAW images has been proposed by Hu et al. [7]. It learns the retouching style from a set of images that takes as reference, and then the style is applied to the RAW source images. The results by this approach are realistic and free of artifacts, but as a deep learning method it needs to be trained with a set of images that represent a certain style. Recently, there is a line of research that increases in popularity: Style transfer using convolutional neural networks [5], [9]. These methods were initially designed to transfer the style of an artwork to a photography [5], and they are able to produce good non-realistic results from images with very different content and style, however they fail in the case of producing photorealistic results free of painting-like distortions. These methods have been adapted to transfer the style between photographies by using a semantic segmentation of the source and reference images [9]. Although the results look visually more satisfying compared to the previous mentioned method, they are not completely free of painting-like artifacts. Instead, we present in this work a method that transfers simple statistic properties from the reference image to the source image, it is a fully automatic approach and does not require any user interaction nor previous training.

The contributions of this work are summarized as follows:

- It directly processes RAW images and it produces a display-ready result that matches the reference look in terms of tone, color-palette and contrast.
- It is a low-complexity method for real-time implementation in-camera.

(a) Source RAW image

(b) Reference image

(c) Luminance non-linear transform
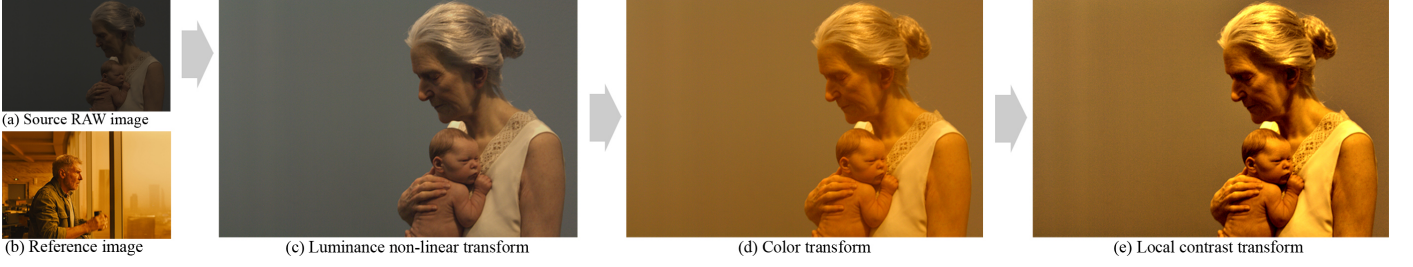
(d) Color transform

(e) Local contrast transform

Fig. 1. Our method provides automatic processing of RAW images, directly from linear RGB data captured by camera sensors to display-ready images. Our system generates appealing results free of artifacts that match the reference style.

- Is a fully automatic method and user interaction or previous training is not required.

II. RELATED WORK

Luminance Tone mapping is a technique to faithfully reproduce high-dynamic-range images in a medium that has a more limited dynamic range. Cyriac et al. [4] presented a tone-mapping method that can accurately reproduce the detail and contrast visible in the original scene captured by the camera. This method is based in natural image statistics and it calculates a variable exponent γ such that when applied to the original (demosaiced white balanced) RAW image I, it performs a constrained histogram equalization. We make an adaptation of this method to make the luminance histogram of the source image move closer to the reference histogram.

Chrominance Since Reinhard et al. [12] presented their pioneering work about color transfer, this has been an active research topic. Their technique takes advantage of the decorrelation property of the $l\alpha\beta$ color space and transfers simple statistical moments (mean and standard deviation) between each channel of the two images, i.e. source and reference. The $l\alpha\beta$ color space is constructed so that it decorrelates natural images, but it cannot be guaranteed to successfully decorrelate others scene types. To overcome this limitation, principal component analysis (PCA) can be used to find a decorrelated color space for each image [8], [14]. When the content of the two images differs, results can be improved by recovering dense pixel correspondence between the images [6], or by having user interaction such as segmenting the images [2].

Contrast Bae et al. [1] proposed an approach to tone management for photographs that can transfer the "pictorial look" from one image to another. Their model decomposes the image into a base and a detail layer with bilateral filtering, and transfers contrast via histogram matching on each layer.

Another approach for contrast adjustment is presented by Cyriac et al. [4] as a second stage in their tone-mapping algorithm. They simulate the contrast normalisation process that happens in the human vision system, in which the contrast (difference between light intensity and its mean value) is divided by a factor depending on the standard deviation of the light intensity. Our method is based on this approach, we adapt it to transfer the local contrast of the reference image to the source image.

III. PHOTOREALISTIC STYLE TRANSFER MODEL

Fig. 1 shows an overview of our method. Given a source RAW image, we seek to generate a photorealistic result that matches the style of the reference image. This is done in three consecutive but separable operations, that can be all performed to have a complete style transfer, or a selection of them to transfer only some of the characteristics of the reference image.

A. Luminance transfer

In this step the demosaiced white balanced RAW image, which is linear RGB data captured by camera sensors, is non-linearly transformed to match the luminance of the reference image.
Using the algorithm proposed by Cyriac et al. [4], that calculates a variable exponent γ such that when applied to the original RAW image I, it performs a constrained histogram equalization creating a non-linear image I_{eq}:

$$I_{eq}(x) = TM(I(x)) = I(x)^{\gamma(I(x))} \tag{1}$$

Let S_0 be the source (demosaiced and white balanced) RAW image and R the reference image. We develop a method based on the above tone-mapping algorithm, that applies a non-linear transformation to the RAW image to match the reference luminance. This transformation is estimated as shown in the following diagram:
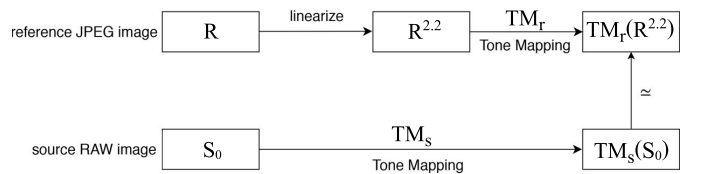

Fig. 2. Estimation of the transformation to be applied to the RAW image so its luminance matches the reference luminance. S is the source image, R is the reference image.

Based on the constrained histogram equalization that the tone-mapping algorithm performs to both the source RAW image and the linearized reference image, we will assume that their luminance histograms are equal:

$$TM_s(S_0) \simeq TM_r(R^{2.2}) \tag{2}$$

So the resulting source image S_1 from this first luminance transfer step will be:

$$S_1 = [TM_r^{-1}(TM_s(S_0))]^{1/2.2} \tag{3}$$

where TM_s and TM_r are the tone-mapping transformations explained in Eq. 1 for the source and reference image respectively and S_0 is the demosaiced white balanced source RAW image.

(a) Reference image R

(b) RAW source image S_0

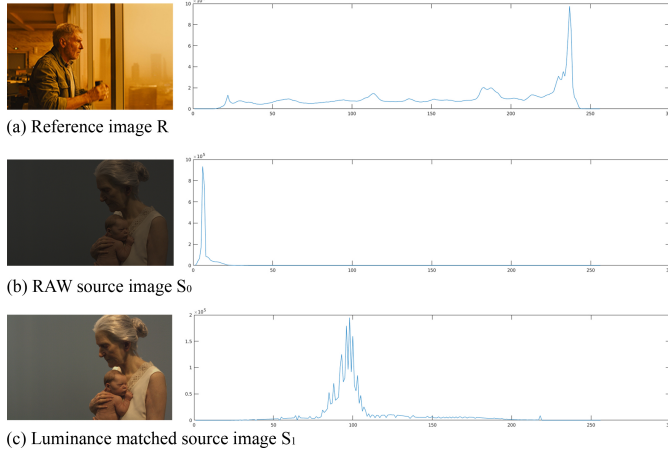(c) Luminance matched source image S_1

Fig. 3. A non-linear transformation is applied to the RAW image so its luminance histogram approximates the reference luminance histogram. The luminance histogram of the source RAW image (b) moves closer to the reference histogram (a), resulting the (c) luminance histogram.

B. Color transfer

(a) Reference image R (b) Luminance matched image S_1 (c) Color transferred image S_2
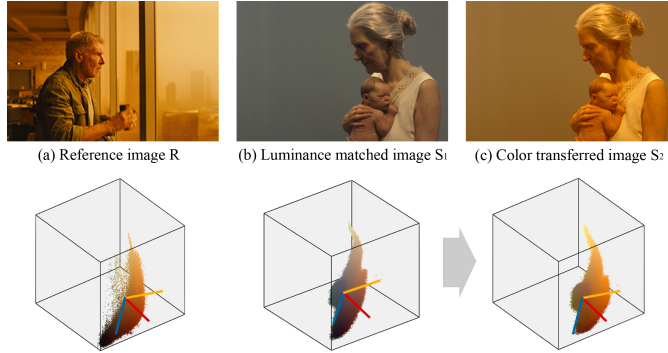
Fig. 4. Geometric interpretation of the transformations made in the color transfer step. Color points shown in RGB reference.

We seek to transfer the colors of the reference image to the resulting non-linear image from the previous step. Statistical properties are often used to define a global mapping between two input images, specially when correspondences between images are not available. Following the same line as in [8], [12], [14], we transfer the statistics (mean and standard deviation) along each channel separately in a decorrelated color space. Principal Component analysis is applied to the RGB source and reference images to find decorrelated spaces for each of them. This method can be summarized as follows: The values of each channel of the source image are first shifted to zero mean by subtracting the mean of the source

from each pixel, they are then rotated to the source PCA decorrelated color space and later scaled by the ratio of standard deviations of both images such that they acquire the standard deviation of the reference image. Finally these values are rotated again to align with the reference PCA decorrelated color space and shifted to match the reference mean. That is a **translation** for matching the mean, a **rotation** for axes alignment and a **scaling** for matching the standard deviation.
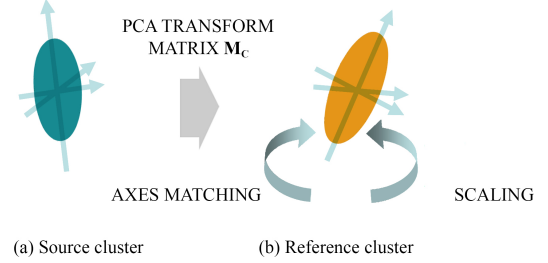

(a) Source cluster (b) Reference cluster

Fig. 5. The statistical properties (represented as an ellipsoid) along the PCA axes of the reference image are transferred to the source image through the matrix M_c.

The matching matrix M_c is given by:

$$M_c = T_{ref} \cdot R_{ref} \cdot V \cdot R_{src}^{-1} \cdot T_{src} \tag{4}$$

where T are the translation matrices, R are the rotation matrices composed of the eigenvectors of the PCA covariance matrices, V is the scaling matrix:

$$V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{\lambda_2^{ref}/\lambda_2^{src}} & 0 \\ 0 & 0 & \sqrt{\lambda_3^{ref}/\lambda_3^{src}} \end{bmatrix}$$

where λ_i are the eigenvalues of the PCA covariance matrices, along the i axis, $i \in \{2, 3\}$.

The resulting source image S_2 from this color transfer step will be:

$$S_2 = M_c \cdot S_1 \tag{5}$$

where M_c is the matrix in Eq. 4 and S_1 is the resulting source image from the previous luminance matching step, obtained in Eq. 3.

This technique has been used in some previous works [8], [14], but we add some modifications to it. As it can be observed in the scaling matrix V, the standard deviation along the first axis of the PCA decomposition is not modified. As it is mentioned in [3], [10], we can assume that the first axis of PCA corresponds to the luminance channel of the image. Luminance is modified in the first and third steps of our method, in this step we only modify the chromaticity channels which correspond to the second and third axes in PCA. The second difference with respect to classic color transfer methods based on PCA decomposition is a constraint added to the rotation matrices to prevent large color shifts. By doing PCA analysis and simply assigning the source axes to the reference axes, the geometry of the resulting mapping might not be as it was intended. The resulting image could have the

color proportions as expected, but locally the colors would be swapped. To avoid that, the method finds the rotation where the source axes are assigned to the closer reference axes.
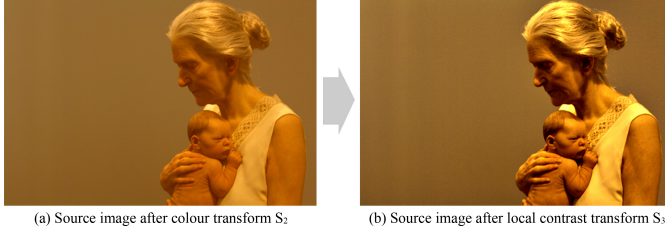
C. Local contrast

(a) Source image after colour transform S_2 (b) Source image after local contrast transform S_3

Fig. 6. Comparison of (a) source image after color transfer step: S_2 and (b) source image after contrast normalization: S_3.

We make an adaptation of the local contrast normalisation formula used in [4]:

$$I'(x) = \mu(x) + (I(x) - \mu(x)) \cdot \frac{k}{\sigma} \qquad (6)$$

where x is a pixel, $I(x)$ is the value at pixel x, μ is the local mean of I, k is a constant, σ is the standard deviation of I, and $I'(x)$ is the final output value for pixel x. The value μ is obtained by convolving I with a kernel W which is a linear combination of two Gaussian kernels.

As we mentioned above, the first component of the PCA keeps the luminance and contrast information of the images, so for modifying the contrast we will work on this first axis of PCA. As a clarification on notation, we will refer to the first PCA channel of an image $I(x)$ as $\bar{I}(x)$.

The resulting image S_3 from this step will be obtained by a local contrast transformation applied to S_2:

$$S_3 = LC(S_2) \qquad (7)$$

where S_2 is the source image obtained from the previous color transfer step in Eq. 5.

We seek to transfer the local contrast from the reference to the source image. If we define local contrast of an image as the difference between light intensity and its local mean value, $I(x) - \mu(x)$, our approach transfers the standard deviation of the local contrast from the reference to the source image through the following formula:

$$\bar{S}_3(x) = \bar{\mu}_{src}(x) + (\bar{S}_2(x) - \bar{\mu}_{src}(x)) \cdot \frac{\sigma_{ref}}{\sigma_{src}} \qquad (8)$$

where x is a pixel, $\bar{S}_2(x)$ is the value at pixel x, $\bar{\mu}_{src}$ is the local mean of $\bar{S}_2$, $\bar{\mu}_{ref}$ is the local mean of R, σ_{ref} is the standard deviation of $\bar{R} - \bar{\mu}_{ref}$, σ_{src} is the standard deviation of $\bar{S}_2 - \bar{\mu}_{src}$ and $\bar{S}_3(x)$ is the output value for pixel x along the first PCA axis of S_2.

The values along the second and third PCA axes of S_2 are not modified in this local contrast transfer step, so S_3 and S_2 are equal along those axes.

(a) Input source and reference images   (b) After local contrast transform
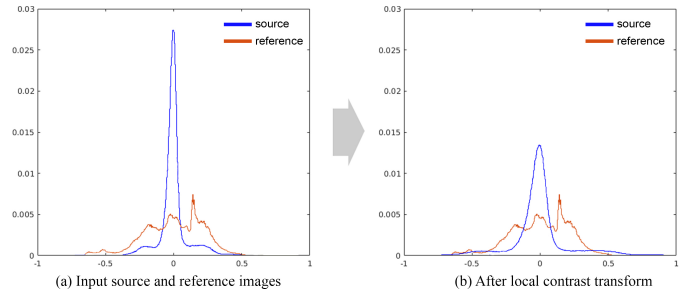
Fig. 7. Comparison of the local contrast histograms before the local contrast transform and after. It can be observed that after the transform the source local contrast histogram is scaled down moving closer to the reference histogram.

IV. RESULTS

The proposed technique has been tested on a number of different scenarios. Some examples are shown in Fig. 8. We demonstrate style transfer results on a wide range of source and reference sequences, where the luminosity and color range vary from one to another. The results are in general visually pleasing, and free of artifacts. However, when the input images are very different in content and composition (as in the 6th row, 4th column of Fig. 8), the results are not as good as expected. A more evident failed result of our method is shown in Fig. 10, it presents color artifacts and the contrast has not been correctly transferred. The main reason for this is the dissimilarity between source and reference images. Generally, statistical color-mapping methods suffer from the same limitation, some of them use segmentation or user interaction to solve this issue. The aim of our method is to ease the workload of cinematographers, during the shooting stage. It can be expected that the selected source and reference images will share some similarities in content and composition: a portrait shooting will be referenced by a portrait image etc. Our method will be suitable for this purpose as not completely different images will be taken as input, producing acceptable results as we have seen in Fig. 8.

The low computational complexity of the approach allows makes makes it running in real-time for 1300×1900 size images. The method requires two histogram calculations for the luminance step, a multiplication by a 4×4 matrix for the color transfer step, and two image convolutions for the contrast transfer stage, therefore it has a very low complexity and is suitable for a real-time implementation. In Fig. 9 we compare our approach to a tone-mapping followed by different color mapping methods. For this purpose, we apply a Naka-Rushton tone-mapping [13] to the source RAW image, this tone-mapping method is fast and simple and it can be compared to ours. Then, to this resulting image, we apply different global, statistical-based color mapping algorithms: Xiao [14], Pitié [11] and Reinhard's methods [12]. Our method is free of artifacts and the preservation of the details in the dark areas is better compared to the other methods.

V. CONCLUSIONS

We have presented a method for real-time in-camera style transfer that produces visually pleasing results free of artifacts

(a) source RAW images (gamma corrected for visualization)

(b) Reference images   (c) Result 1 images   (d) Result 2 images   (e) Result 3 images
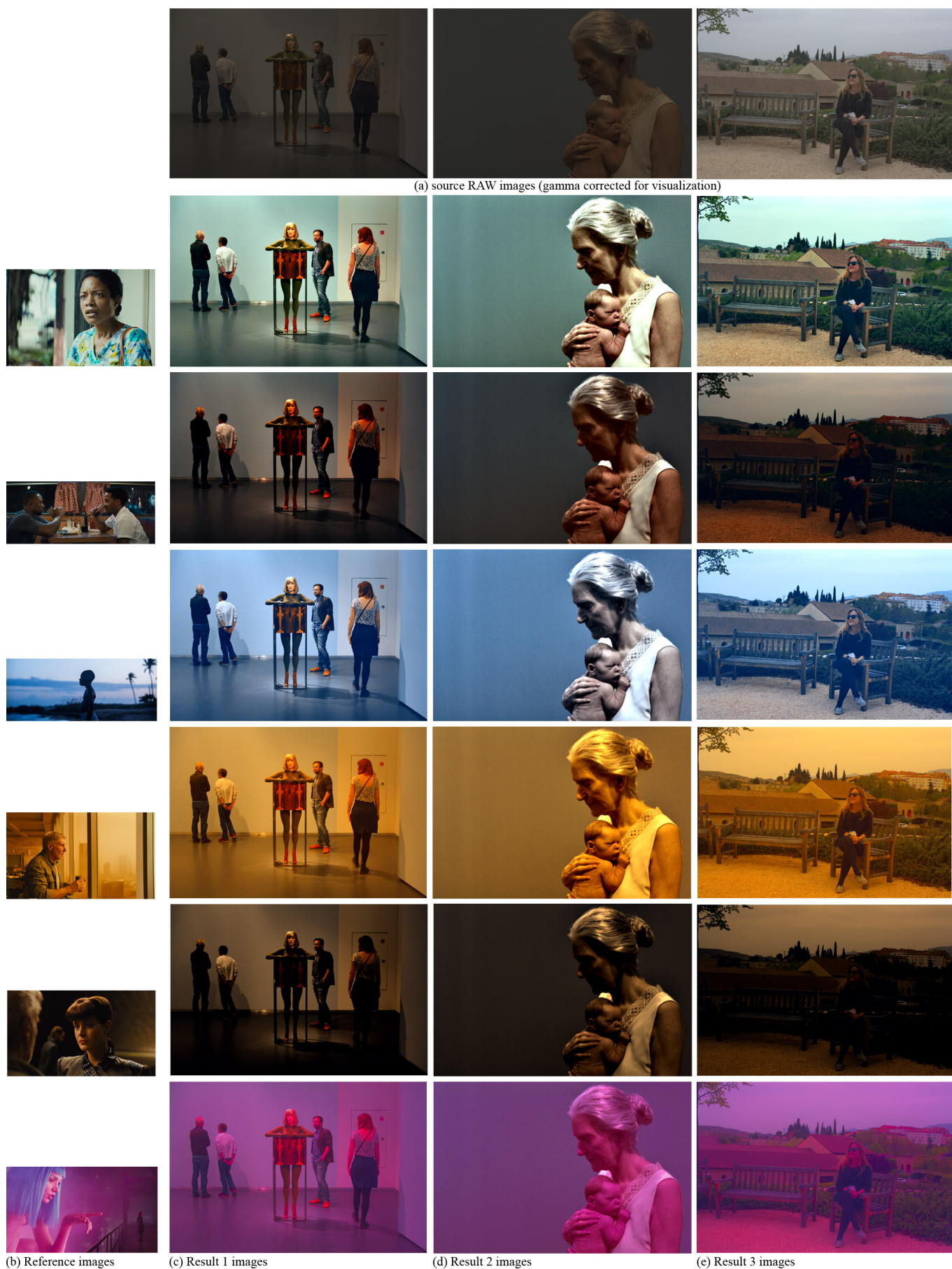
Fig. 8. Our method can successfully transfer a variety of styles to a range of source images. Here we demonstrate this for different styles characterized by scenes from movies graded by professional colorists: 'Moonlight' and 'Blade Runner 2049'.

(a) Source image (gamma corrected for visualization)

(c) Naka-Rushton tone mapping + Xiao method

(e) Naka-Rushton tone mapping + Reinhard method

(b) Reference image

(d) Naka-Rushton tone mapping + Pitié method

(f) Our method

Fig. 9. A comparison of our model to some well-known color-mapping methods. (a) and (b) are the source and reference images respectively, (c), (d) and (e) images are obtained by applying a Naka-Rushton tone-mapping method to the source RAW image, followed by different global, statistical color mapping techniques: Xiao, Pitié and Reinhard's methods.

(a) Source RAW image (scaled for visualisation)

(b) Result of our method

(c) Reference image

Fig. 10. The difference in content and dynamic range of the input images causes color artifacts and a non natural looking result.

that look photorealistic even for challenging scenes. It is based in basic statistical properties of the input images and is able to transfer the global style in very different scenarios. As future work we aim to implement our method for video sequences. The naive approach would be to apply our method frame by frame to the video, that might result in creating artifacts such as color bleeding and temporal incoherence. For avoiding that, some temporal coherence has to be incorporated to the results given by applying our method to each frame of the video sequence.

VI. ACKNOWLEDGMENT

REFERENCES

[1] S. Bae, S. Paris, F. Durand, Two-scale tone management for photographic look, ACM Transactions on Graphics (Proceedings SIGGRAPH 2006), 2006.

[2] N. Bonneel, K. Sunkavalli, S. Paris, H. Pfister, Example-based video color grading, ACM Transactions on Graphics (Proceedings of SIGGRAPH 2013), 2013.

[3] M. Brown, S. Susstrunk, Multi-spectral SIFT for Scene Category Recognition, International Conference on Computer Vision and Pattern Recognition , 2011.

[4] P. Cyriac, D. Kane and M. Bertalmío, Optimized tone curve for in-camera image processing, IS&T Electronic Imaging Conference, 2016.

[5] L. A. Gatys, A. S. Ecker, M. Bethge, Image style transfer using convolutional neural networks, IEEE Computer Vision and Pattern Recognition, 2016.

[6] Y. HaCohen, E. Shechtman, D. B. Goldman, D. Lischinski, Non-rigid dense correspondence with applications for image enhancement, ACM Transactions on Graphics (Proceedings of SIGGRAPH 2011), 2011.

[7] Y. Hu, H. He, C. Xu, B. Wang, S. Lin, Exposure: A white-box photo post-processing framework, ACM Transactions on Graphics SIGGRAPH, 2018.

[8] H. Kotera, A Scene-referred color transfer for pleasant imaging on display, IEEE International Conference on Image Processing, 2005.

[9] F. Luan, S. Paris, E. Shechtman, K. Bala, Deep photo style transfer, IEEE Computer Vision and Pattern Recognition, 2017.

[10] M. Mudrov, A. Prochazka, Principal Component Analysis in image processing, Proceedings of MATLAB Technical Computing Conference, 2005.

[11] F. Pitie, A. Kokaram, R. Dahyot, Automated colour grading using colour distribution transfer, Computer Vision and Image Understanding, 2007.

[12] E. Reinhard, M. Ashikhmin, B. Gooch and P. Shirley, Color transfer between images, IEEE Computer Graphics and Applications, 2001.

[13] E. Reinhard, K. Devlin, Dynamic range reduction inspired by photoreceptor physiology, IEEE transactions on visualization and computer graphics, 2005.

[14] X. Xiao, L. Ma, Color transfer in correlated color space, VRCIA, 2006.